

## CA3 - Group 1

### 1

*Try to solve this optimization task with proper choices of size of decision variables (matrix  $W1$ , matrix  $W2$ , and vector  $w3$ ) using GD, perturbed GD, SGD, SVRG, and block coordinate descent. For the SGD method, you may use the mini-batch version*

- They implemented feed-forward and backward propagation, although in a different way. GD, SGD, and SVRG algorithms are also implemented. They did not implement PGD and BCD.
- They trained the algorithms on a subset of samples from the "Individual household electric power consumption" dataset and compared their training times. The results showed that SGD was the fastest, although the figures for loss did not appear to be accurate.
- It would be better if the authors had plotted all the results in one figure instead of separate ones for each algorithm.
- There was no information provided on the analysis of the "Greenhouse Gas Observing Network" dataset or whether the algorithms were tested on it.

### 2

*Compare these solvers in terms complexity of hyper-parameter tuning, convergence time, convergence rate (in terms of number of outer-loop iterations), and memory requirement*

- They did not answer to this question explicitly. However, they aimed to show the running time of each algorithm, although the figures are not correct.